Model of cognitive processes and conversational principles in survey interview interaction

Ongena, Y.P.; Dijkstra, W.

published in
Applied Cognitive Psychology
2007

DOI (link to publisher)
10.1002/acp.1334

document version
Publisher's PDF, also known as Version of record

APPLIED COGNITIVE PSYCHOLOGY
Appl. Cognit. Psychol. **21**: 145–163 (2007)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/acp.1334

A Model of Cognitive Processes and Conversational Principles in Survey Interview Interaction

YFKE P. ONGENA[1,2]* and WIL DIJKSTRA[1]

[1]*Vrije Universiteit Amsterdam, The Netherlands*
[2]*University of Nebraska-Lincoln, USA*

SUMMARY

In this paper we provide a model of interviewer–respondent interaction in survey interviews. Our model is primarily focused on the occurrence of problems within this interaction that seem likely to affect data quality. Both conversational principles and cognitive processes, especially where they do not match the requirements of the respondent's task, are assumed to affect the course of interactions. The cognitive processes involved in answering a survey question are usually described by means of four steps: interpretation, retrieval, judgement and formatting. Each of these steps may be responsible for different overt problems, such as requests for clarification or inadequate answers. Such problems are likely to affect the course of the interaction through conversational principles which may cause, for example, suggestive behaviour on the part of the interviewer, which may in turn yield new problematic behaviours. However, the respondent may not be the only one who experiences cognitive problems; the interviewer may also have such problems, for example with respect to explaining question meaning to the respondent. Thus the model proposed here, unlike most of the other models which concentrate on the respondent, tries to incorporate cognitive processes and conversational principles with respect to both interviewer and respondent. In particular, the model looks at how cognitive processes and conversational principles affect both the interaction between interview participants and the quality of the eventual answers. Copyright © 2007 John Wiley & Sons, Ltd.

In survey interviews the interviewer performs multiple tasks like asking questions, evaluating answers, probing, motivating respondents and recording the eventual answer. Respondents have only one task: answering the questions. The tasks of both interviewer and respondent can be described from a cognitive viewpoint, applying cognitive theories about information processing and memory. This cognitive approach has been primarily used to describe the respondent's part of the question–answer process. It is assumed that answering a survey question involves several cognitive steps as described in the well known four-step model of survey response, that is interpreting the question, retrieving relevant information from memory, forming a judgement from the retrieved information and formatting the response (see Tourangeau, Rips, & Rasinski, 2000).

A striking characteristic of this cognitive model of survey response, and most of the other models cited by Tourangeau et al. (2000), is that the interviewer is not included as a

*Correspondence to: Yfke P. Ongena, Department of Social Research Methodology, Vrije Universiteit Amsterdam, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands. E-mail: yp.ongena@fsw.vu.nl

factor. We think that to better understand why and when problems occur in the interaction, the interviewer should be included in the model.

Sander, Conrad, Mullin, and Herrman (1992) mention the need for three types of models that account for the mental processes of the interviewer: 'The interviewer model of question generation', 'The interviewer model of question clarification' and 'The interviewer-respondent interaction model'. They present instances of the first two model types but unfortunately, they do not present an interaction model. This is not surprising as the interaction between the interviewer and respondent may be too complex to be clearly modelled.

Yet it may be this very complexity that shapes how respondents produce answers. As a simple example, if the respondent has problems in understanding the meaning of the question, he may ask for clarification. The interviewer may provide such clarification, or, adhering to the rules of standardised interviewing (Fowler & Mangione, 1990), may respond with 'Whatever it means to you'. While certainly not leading, 'Whatever it means to you' probes may be dangerous because they leave the interpretation of the question up to the respondent—in fact they instruct the respondent not to speculate about the intended meaning of the question—which could well lead to an incorrect interpretation and increase the chances of an erroneous response (Schober & Conrad, 2002). In addition, because ordinary words may have special or technical meanings in survey, whether or not respondents obtain clarification may have a profound impact on the accuracy of their responses; clarification in form of definitions can make all the difference between answering correctly and incorrectly (Conrad & Schober, 2000; Schober & Conrad, 1997). Moreover, particular conversational principles may affect whether or not the respondent will request for clarification at all. Several studies have pointed out that in survey interviews, interviewers and respondents communicate according to rules of ordinary conversations. Thus, in order to understand more fully the course of the interaction between interviewer and respondent, conversational principles must also be taken into account.

In this paper we will present a comprehensive model of interviewer-respondent interaction in survey-interviews. The model is primarily directed towards communication problems that we suspect affect the quality of an eventual answer, or at least to be related to that quality. Unlike other models we are aware of, our model incorporates both interviewer and respondent behaviour as well as both cognitive and conversational aspects. We partition our discussion of the model into three separate sections. The first concerns the interviewer's cognitive processes involved in asking the questions, and is an adaptation of Sander et al.'s 'interviewer model of question generation'. The second part concerns problems that respondents may have in understanding survey questions. In the third part, cognitive problems of respondents with respect to retrieval, judgement and formatting are presented.

INTERVIEWER'S QUESTION FORMULATION

The verbal processes in survey interviews take place in so-called question–answer sequences (Q–A sequences). Such a sequence starts with the interviewer asking a question, and it ends when the next question is posed, indicating that the interviewer has acknowledged the respondent's answer (Dijkstra, 1999). See Excerpt 1 for an example:

DOI: 10.1002/acp

Excerpt 1: example of a Q–A sequence

1. I: And uh what percentage of the time did you then watch with real attention when the TV was on?
2. R: Well, uh forty I guess
3. I: Forty

This Q–A sequence was taken from a Television Survey (Smit & Neyens, 2000) and translated from Dutch.

In Line 1 the interviewer poses the question. The respondent gives an answer in Line 2 which is repeated by the interviewer to show that she acknowledges the answer.

In order to get reliable data, a fundamental assumption in standardised interviewing is that interviewers must read all the questions exactly as worded (Fowler & Mangione, 1990). In practice, however, interviewers may not read questions exactly as worded for a variety of reasons. In ordinary conversations, utterances are adapted to specific recipients, for example children versus adults ('recipient design', Houtkoop-Steenstra, 2000; Suchman & Jordan, 1990). In survey-interviews, however, question wording is determined in advance, and usually designed for a large and heterogeneous group of recipients, adapted to all possible circumstances ('audience design', Houtkoop-Steenstra, 2000). As Suchman and Jordan (1990) argue, scripting questions for multiple groups of respondents often results in long and awkwardly structured questions that are difficult to read. Furthermore, for questionnaire designers it is impossible to account for all possible circumstances. Thus, interviewers have the tendency to breach standardisation rules, and read questions in their own adapted wording (Suchman & Jordan, 1990).

The need for systematically evaluating interviewer performance (and other behaviours in interviewer-respondent interaction) was recognised almost 40 years ago by Cannell and associates (Cannell, Fowler, & Marquis, 1968; Marquis & Cannell, 1969), when they designed the first behaviour coding schemes. In a review of 26 studies that evaluated question reading (Ongena & Dijkstra, 2006), the percentage of questions read exactly as worded varied from 28 to 97 of the questions across different surveys. For survey researchers it is interesting to know more about the causes and effects of such variations in question reading.

Sander et al.'s 'interviewer model of question generation' is a first attempt to model the cognitive processes involved in reading questions. We have slightly adapted and elaborated this model (Figure 1). Our model, like Sander et al.'s, concerns completion of *one* question within an interview. The number of times an interviewer has conducted the same interview and posed the same questions in previous interviews (a) affects the availability of question wordings in the interviewer's memory (b).[1] When the question wording is available, it is possible that the interviewer may glance at only some of the question words, the question number, or the specific page or screen of the questionnaire (d), otherwise she has to read the question text (c). When interviewers are committed to verbatim reading, they encode the question words (p), or reproduce the question from memory (q). When interviewers are not committed to verbatim reading they may interpret (n) or reconstruct (o) the question's

[1]Parenthesized letters in the text refer to the corresponding component of the model in the figure.

DOI: 10.1002/acp

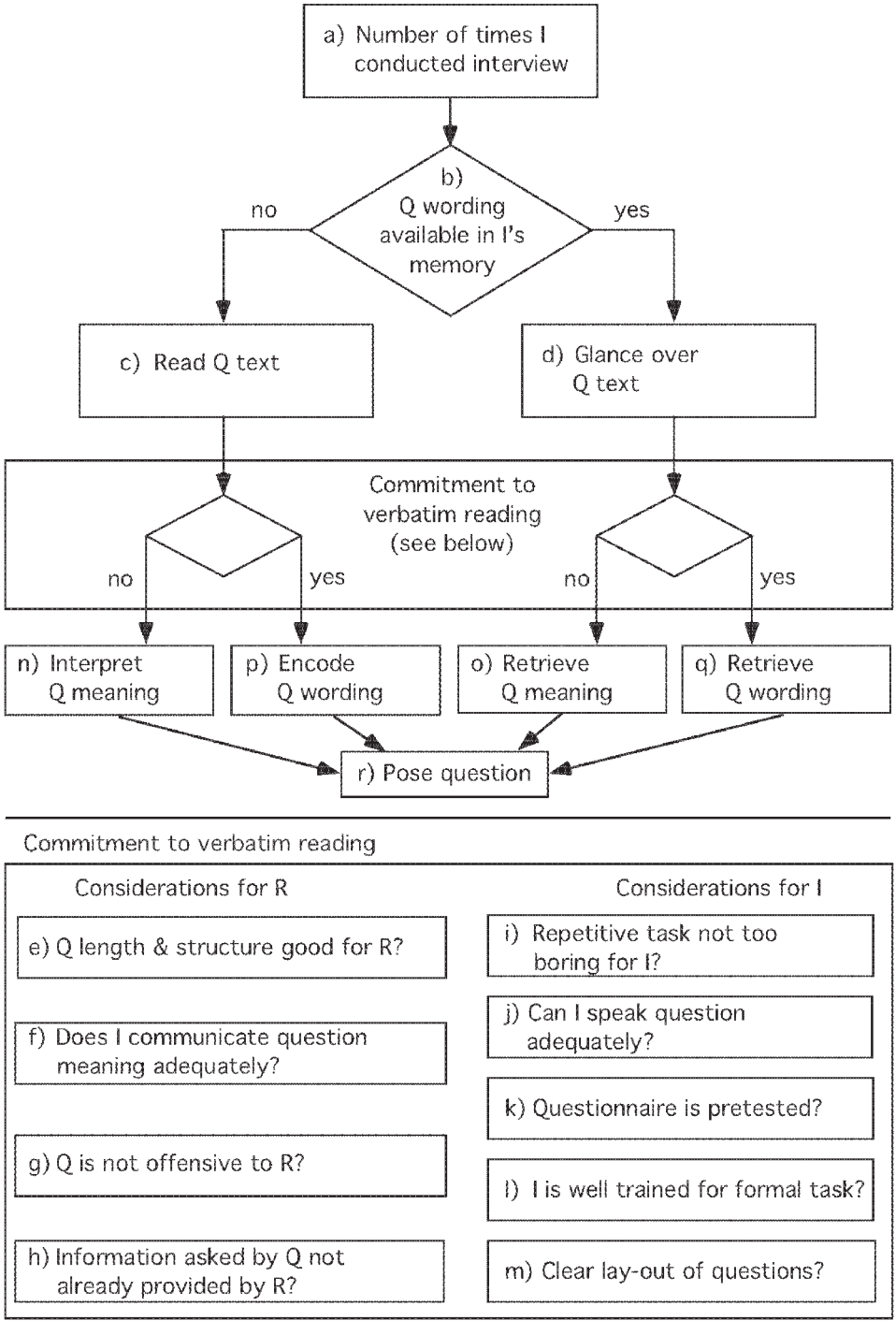

Figure 1. Model of interviewer's question formulation (partly based on Sander et al., 1992)

DOI: 10.1002/acp

meaning to pose the question (r). We expect interviewers to rely more on their memory for the question the more times they have asked it. Note also that a verbatim presentation of the question is possible without visually reviewing all the words in the question if the interviewer's memory for the question is sufficiently complete.

As Sander et al. argue, many deviations in question reading concern more or less deliberate paraphrases of the original question. Interviewers may paraphrase questions when the original question is lengthy, awkwardly structured or just difficult to read, as judged for themselves (j) and for respondents (e). As Bradburn and Sudman (1980) point out, when question length increases, errors and variance in question reading are likely to increase. Cahalan, Mitchell, Gray, Chen, and Tsapogas (1994) also found that long questions yielded more problematic interviewer behaviours, such as variations in question reading, especially for questions with qualifying statements. An example of such a question, taken from the 2002 Dutch pilot of the European Social Survey is:

> 'Regardless of your basic or contracted hours, how many hours do/did you *normally work* a week (in your main job), including any paid or unpaid overtime'.

The question was not read as worded in 38% of the cases; the interviewers never included the last qualifying statement ('including any paid or unpaid overtime'), and also found other ways to decrease question length (Ongena, 2003). Such omissions are very likely to affect question meaning and thus will affect data quality in a negative way.

Although theories of involvement (Hyman, Cobb, Feldman, Hart, & Stember, 1954) and satisficing (Krosnick, 1999) are generally focused on respondent behaviour, we think these principles are also important in explaining interviewer behaviour, that is omissions may be a matter of low task involvement of the interviewer. However, Dykema, Lepkowski, and Blixt (1997, p. 304) found that, interviewers' adapting question wording to specific situations, by omitting potentially inapplicable or previously mentioned parts of a question, was positively (although for most questions not significantly) related to response accuracy. Such variations in question wording may very well be caused by a *high* level of task involvement. Hence, the eventual effect on data quality may depend on the reason for the interviewer's rewording: to ease question reading for the interviewer or to make the question easier for the respondent. If interviewers judge questions as difficult to be understood (f), they may also choose to add definitions of concepts judged as difficult for certain respondents. Research by Conrad and Schober (2000) showed that presenting definitions enhanced the overall response accuracy, even though this practice led to different question stimuli for different respondents. No research has been done yet to examine the ability of interviewers to adequately judge respondents' potential problems and to translate this judgement into adequate question wording, although interviewers in the Schober and Bloom (2004) analyses seemed to effectively discriminate those cues where clarification was helpful from those where it was not.

Interviewers may also paraphrase questions they believe will be offensive (g) or might otherwise impair establishing rapport with the respondent. Interviewers may reduce the negative face, allowing the respondent to have more freedom to act as they choose (i.e. leaving more room for freedom of acting to the listener, Brown & Levinson, 1987), by modifying the question by rewording. An interviewer may for example spontaneously add a phrase like 'Could you tell me' to a question like 'What was your monthly income during the past 12 months?' This rewording gives the respondent more freedom by making it acceptable to not answer the question. However, such strategies often lengthen the question

DOI: 10.1002/acp

and thus may reduce their clarity. Moreover, as in the example above, item non-response may be higher, if the respondent chooses to not answer the question. As far as we know, no research has been conducted yet concerning how rewording questions to make them more polite affects data quality.

Interviewers may also adjust question wording to account for information that the respondent already spontaneously provided (h). For example, a respondent may have already reported that he is retired. When an interviewer next needs to ask the respondent's current employment status, adherence to exact question wordings may be awkward. Hence, interviewers often accompany a redundant question with a provisional answer, or add remarks like 'You've already said it but I have to ask'. Such 'self-repairs' indicate that the question is 'retrospectively redefined by the interviewer as a case of reading a scripted line' (Houtkoop-Steenstra, 2000, p. 77). Thus, the interviewer signals the respondent that she is reading 'audience-designed' questions, which may decrease the respondents' motivation to give accurate answers relative to a question designed (tailored) specifically for the respondent. Respondents' answers may also directly be affected by such self-repairs. For example, the information spontaneously provided (e.g. 'being retired') does not necessarily have to match the question's definition (e.g. 'being unemployed'). Nevertheless, by stating 'you have already said it' the interviewer indicates that the information does match the definition. To avoid such negative effects on data quality, 'self-repair' should be formulated as neutrally as possible (e.g. 'We already talked about it' instead of 'You already told me you are retired').

Interviewers may also solve the interactional problem by not posing the redundant questions at all, recording the information provided earlier by the respondent as if it were an answer to the unasked question. In that case respondents are not given a chance to correct possibly invalid inferences based on the information provided earlier. Ongena and Dijkstra (2006) observed that up to 22% of all the questions in a survey may have been incorrectly skipped. While this almost certainly reflects some general strategies to decrease the interview time at least some of such omissions are caused by entering already provided information or inferences based on this information.

The repetitive character of survey questionnaires may cause boredom (i), discouraging interviewers from reading questions exactly as worded (Mathiowetz & Cannel, 1980). Repetitive scripts may also discourage respondents (Suchman & Jordan, 1990). To our knowledge, though, the effects of repetitive scripts on data quality have never actually been studied.

Reading questions exactly as worded may be easier when question wordings are carefully pretested (k), and when interviewers are well trained and specifically instructed to read questions as worded (l). Research has shown that interviewers who did not receive training perform worse than trained interviewers (Billiet & Loosveldt, 1988). Furthermore, it may be useful to provide interviewers with clearly presented questions (m), for example using large font and a layout that clearly distinguishes questions from instruction texts. Pierzchala and Manners (1998) suggest that interviewer performance can be improved when question and screen formatting conventions are used, especially for multiple surveys within a single programme. For more specific details on screen design issues see Hansen and Couper (2004).

In addition options in CAI-software can be used to automate routine tasks, which allow interviewers to pay more attention to other tasks (Sperry, Edwards, Dulaney, & Potter, 1998, p. 364–365). Question wording can be adapted to respondents' answers to earlier questions. In the pilot study of the European Social Survey, interviewers were asked to

DOI: 10.1002/acp

reword work related questions in either present or past tense, depending on the previously asked respondent's work status. Apparently, this instruction was difficult to apply; in 24–43% of the instances interviewers actually did not reword the questions (Ongena, 2003). Thus, to adapt question wordings to respondents' specific situations, it is better to let the CAI software do this task.

RESPONDENT'S INTERPRETATION OF A QUESTION

After the interviewer has read the question, some cognitive processing is required by the respondent to understand the question. If the respondents do not uniformly understand particular questions as they are intended by the researcher, responses will not be comparable because different respondents will answer what are essentially different questions.

As Sudman, Bradburn, and Schwarz (1996) suggest, for interpretation of questions, a distinction can be made between understanding the literal and the pragmatic meaning of a question. Respondents may have trouble understanding the literal meaning as a consequence of lexical or structural ambiguities. Lexical ambiguities such as a foreign or unconventional word may be resolved by providing clarification. A repetition may be sufficient to solve a problem of structural ambiguity, for example an odd word order in the question.

When respondents have no problems in understanding the literal meaning of a question, they may still have problems understanding its pragmatic meaning. For example, in the question 'How many days a week do you use butter?' no uncommon words or uncommon sentence structure can be found. However, a respondent may incorrectly interpret butter to include margarine. Unless the respondent requests clarification of the pragmatic meaning of 'butter' such a misunderstanding may go unnoticed. This will then result in an over-reporting of butter consumption.

Figure 2 shows a model of respondent and interviewer behaviour related to question comprehension. The first relation in this model implies that question characteristics (e.g. ambiguity, presence of complex concepts or unfamiliar words) will influence (i.e. degrade) question comprehension (b). The relations b–c versus d/e/f comprise how respondents deal with problems in understanding.

In ordinary conversations, participants will try to solve problems, by initiating 'repair' (i.e. clarifying ambiguous constructs) before they continue the ongoing conversation (Schober, 1999). In a survey interview, a respondent may perform actions that signal problems in understanding either explicitly (d), implicitly (e) or not at all (f). Consider the following different respondent reactions to the question 'How many days a week do you use butter?'

R1: What do you mean by butter?
R2: Does margarine count as butter?
R3: I like to use margarine every day
R4: Almost everyday
R5: That is a difficult question
R6: 5 days a week

Explicit requests (d) may be requests for clarification (g, example R1), requests to repeat the question (h) or clarification 'proffers' (i). The latter are utterances like example R2 in

DOI: 10.1002/acp

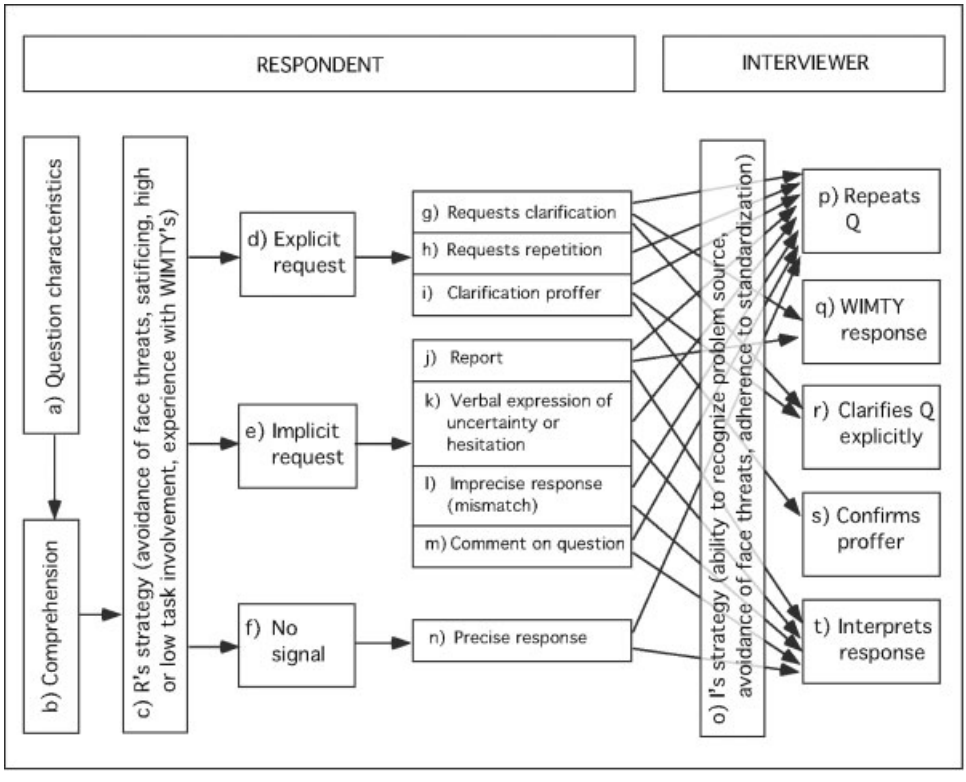

Figure 2. Model of respondent and interviewer behaviour related to question comprehension

which the respondent exhibits some question interpretation (Moore & Maynard, 2002). Explicit requests for clarification are relatively rare in survey interviews (Schaeffer & Maynard, 2002, p. 293) and hardly ever occur in strictly standardised interviews. In a corpus of interviews in which respondents answered questions based upon fictional scenarios, they explicitly requested clarification in 2% of the questions in complicated scenarios, and never in straightforward scenarios (Schober & Bloom, 2004, p. 296). Ongena and Dijkstra (2006) report that such requests occurred in at most 23% of the questions in surveys they examined in their review. Respondents may be discouraged to explicitly request clarification, as they know the response will be a standardised phrase like 'Whatever it means to you' (a 'WIMTY' response, Moore, 2004). Schober, Conrad, and Fricker (2004) actually found that when interviewers are allowed to clarify questions, that is not providing 'WIMTY' responses, respondents are more likely to request clarification.

Actions that mark but do not specifically address respondents' problems, that is implicit requests (e), are more common. Example R3 is a 'report' (j), that is a description of the respondent's situation rather than a response. By giving reports, respondents leave the interpretative judgement involved in answering the question (e.g. is margarine considered butter?) to the interviewer (Moore, 2004) and when they are particularly frequent for certain questions, reports can be a signal that the question needs revision (Schaeffer & Maynard, 2002). A report is also a more efficient face-keeping strategy than providing an actual response; respondents do not explicitly request clarification, nor do they

DOI: 10.1002/acp

acknowledge that they are having trouble understanding, whereas a request for clarification does both.

Other signs of respondent problems are hesitations (k), imprecise answers ((l), example R4 above) or commenting on the question ((m), example R5), which can be interpreted as a 'don't know' answer or an implicit request for clarification. Finally, respondents can also give precise answers ((n), example R6), without any indication of problems at all (f). Presumably it is easier for interviewers to recognise explicit than implicit requests, which are in turn easier to recognise than seemingly trouble-free answers when the respondent does not realise that he or she misunderstands the question.

We hypothesise that the likelihood of explicit requests is increased by a high task involvement, a low level of experience with WIMTY's ('Whatever it means to you'), a high level of experience with useful clarifications and no tendency to avoid negative face-threatening actions. The likelihood of the various forms of implicit requests on the other hand, is enhanced by frequent WIMTY's, a low level of task involvement and a high tendency to avoid negative face threats.

Interviewers have several options for clarifying the question meaning in response to signs of uncertainty. First, they have the option—and under some interviewing approaches they are required—to *not* clarify question meaning. But if they are able or willing to clarify the question, they are more likely to do so the more explicit the signal (explicitness varies across respondents' utterances (g–n)) primarily because they are more likely to notice it.

If they do respond, the interviewers' reaction will depend on their commitment to standardisation rules versus conversational principles (o). According to standardisation rules (Fowler & Mangione, 1990), interviewers are only allowed to repeat questions (p) or give a 'WIMTY' response (q). When interviewers do not (have to) adhere to such rules: they can clarify the question (r), or confirm a proffer (s). From analyses of interview interactions, Moore and Maynard (2002, p. 296) concluded that interviewers were twice as likely to respond in an unstandardised way to clarification proffers than to explicit requests for clarification. Clarification proffers (such as 'Does margarine count as butter?') not only indicate the source of the problem in understanding (e.g. the definition of butter) but also comprise an offer of candidate clarification (e.g. 'If the definition of butter includes margarine then I will consider margarine in my response'). When respondents use these clarification proffers, it is very easy for interviewers to change the meaning of the question, because all that is needed is a short acknowledgement (Moore & Maynard, 2002). In the butter question example a standardised reply, that is no clarification, may lead to over-reporting of butter consumption.

Reports usually enable the interviewer to infer what the answer should be (t). As Moore argues, with these reports, respondents avoid negotiating the problem explicitly. This makes them more efficient (from the viewpoint of the respondent) than requests for clarification. The result may be that standardised interviewers, having learned to not clarify survey questions, may be more likely to infer an answer, rather than to actively and collaboratively negotiate the judgement (although this is an egregious violation of standardised practice). Whether they choose to do this will depend on their commitment to and understanding of standardisation rules.

The avoidance of face threats may also play a role in how the interviewer reacts to respondents' reports; face threats may be avoided by simply inferring a codeable answer from the respondent's report without pressing the respondent for an explicit answer. After all, respondents have done their best to arrive at an answer, so probing for an explicit answer may bother them. Finally, time constraints may also prompt the interviewer to infer

DOI: 10.1002/acp

the answer from a report because this is the quickest approach. Such inferences are likely to affect data quality negatively. For instance, it would be inadequate to infer from example R3 ('I like to use margarine every day') that the respondent does not use butter.

The reason for not allowing interviewers to clarify question meaning originates from a standpoint of offering standardised *stimuli* to respondents. However, such standardisation of stimuli does not necessarily mean standardisation of *meaning*. These considerations of clarification of ambiguous concepts were the basis of Conrad and Schober's studies (Conrad & Schober, 2000; Schober & Conrad, 1997, 2002; Schober et al., 2004). They experimentally compared standardised interviewing techniques with conversational interviews, where interviewers were able to provide clarification. Response accuracy appeared to be higher in conversational interviews than in strictly standardised interviews, particularly when question concepts were unclear without definitions. However, this result came at a price; conversational interviews took longer than standardised interviews in proportion to the amount of clarification (Schober et al., 2004, Experiment 1). The success of conversational interviewing also depends on the extent to which interviewers are able to recognise implicit requests for clarifications as such. As Conrad and Schober (2000) indicate, 96% of the clarifications were given when it was not explicitly asked for. In some cases the clarification was prompted by a request to repeat the question; however, in most of the cases clarifications were given even though the respondent did not give explicit signs that they needed clarification.

RETRIEVAL AND JUDGEMENT

In Figure 3 the remaining parts of the model are shown. In this section, processes related to retrieval and judgement of information will be described. Formatting the response will be described in the next section.

After the respondent and interviewer have dealt with any comprehension difficulties (a1, more fully illustrated in Figure 2), a respondent may verbally express retrieval and judgement (b–b1). Tourangeau et al. (2000, p. 146) give a clear overview of retrieval strategies that respondents may adopt when answering behavioural frequency questions. Two of those strategies are not likely to be visible in the interaction; that is when respondents have exact tallies available or when they provide direct estimations based upon general impressions. When respondents use one of those strategies, they are likely to immediately produce an 'acceptable' answer that can be recorded by the interviewer without further interaction.

The kind of strategy that survey researchers often hope for (or even assume), is that respondents recall each and every relevant event, and enumerate all events to get their answer ('recall-and-count' or 'episodic enumeration'). Although not required to do so, respondents may start to verbally express their enumeration, as happens in Excerpt 2, Line 2. Respondents may start such verbal expressions to show the interviewer that they are busy with processing. Another reason for these expressions may be that it is just helpful to think aloud while retrieving relevant information, or it may be a way to elicit help from the interviewer. Whatever the reason for their occurrence, these expressions may trigger the interviewer to interrupt, offering some help, or even inferring the respondent's answer. Lines 3 and 5 in Excerpt 2 show such an interruption. The interviewer translates the respondent's story into an estimation of total viewing time to obtain a codeable response. The interviewer suggests an answer instead of stimulating the respondent to calculate the

DOI: 10.1002/acp

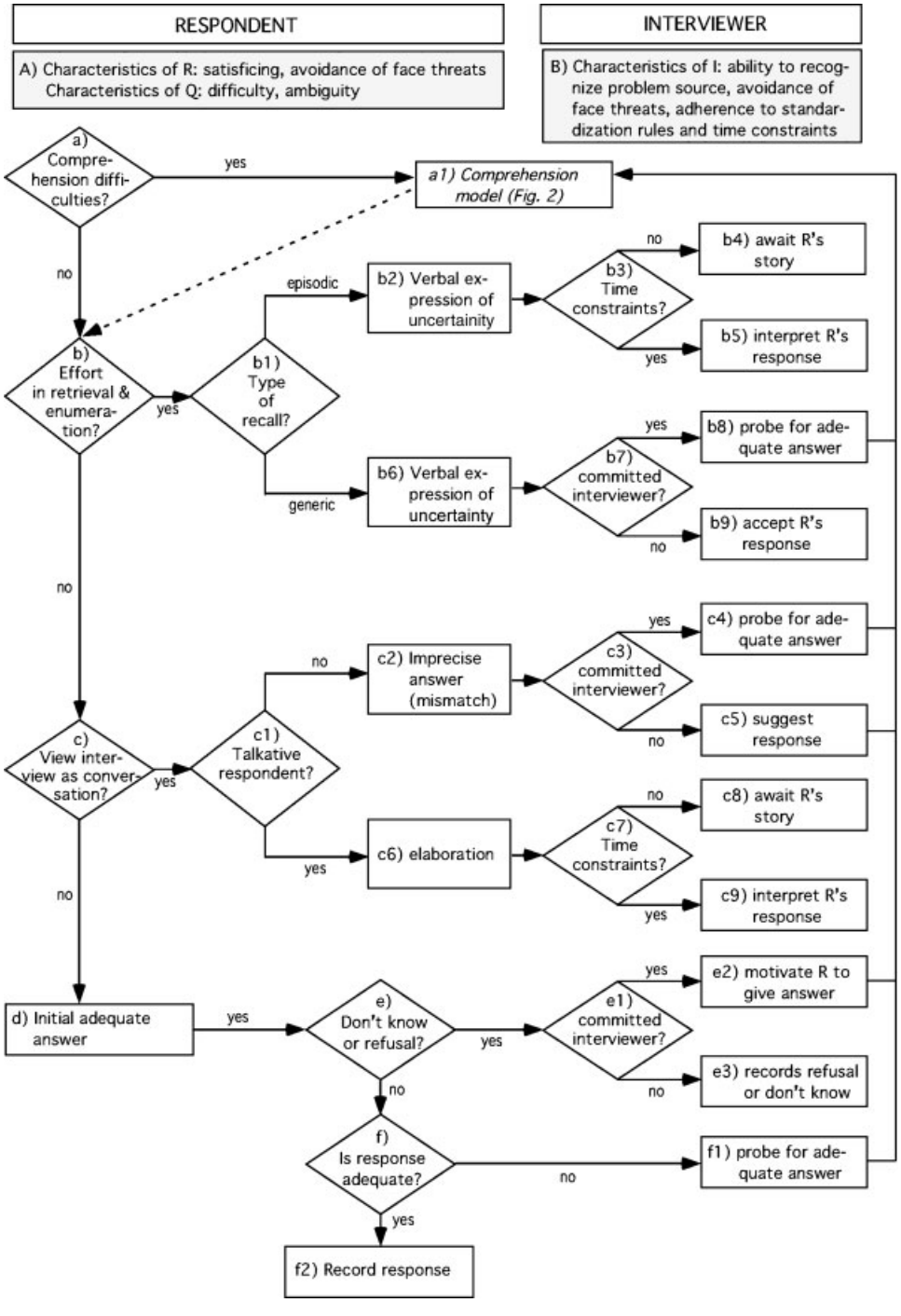

Figure 3. Model of respondent-interviewer interaction in the retrieval, judgement and formatting phases

DOI: 10.1002/acp

exact number of hours and minutes based on what was reported in Line 2. Interrupting the enumeration process will likely result in underreporting, as is probably the case in this example.

Excerpt 2: verbal expression of enumeration in the interaction

1. I: And uh how many hours or minutes did you watch television? So just.
2. R: Oh well yes. Think Tank that started at seven thirty and then the news until twenty past eight and then I got uh guests at the door . . .
3. I: Yeah so uh . . .
4. R: Then I turned it off
5. I: . . . about an hour.
6. R: Yes
7. I: Okay

This Q–A sequence was taken from a Television Survey (Smit & Neyens, 2000) and translated from Dutch.

Generally, because of the time constraint in which they usually have to operate (b3), interviewers are likely to cut-off respondent's stories (b5), instead of awaiting the respondent's full enumeration (b4). However, stories may also be useful, especially when complex events are being retrieved. A study by Means, Swan, Jobe, and Esposito (1991) showed that asking questions produce 'event stories' or 'narrative forms' may facilitate respondents to reconstruct relevant information to answer retrospective questions. This information may then be used to give more accurate answers.

Although these examples refer to questions about autobiographical information, similar processes may occur when respondents answer attitude questions. When respondents are asked their opinion on complex matters (e.g. their opinion towards the government) they may try to retrieve relevant information (their opinion on current governmental issues, political parties, members of the parliament, etc.) and express their considerations to the interviewer during this retrieval. No research has been done yet to explore how often and in what circumstances respondents spontaneously and overtly enumerate in response to closed-ended autobiographical or attitude questions, and how frequently interviewers interrupt such enumeration. Again, interruptions may degrade data quality as the respondent is not given time to take all relevant factors into account.

Respondents may also use generic information to form their judgement. For example, respondents may retrieve information about the rate of occurrence of events or behaviours, without recalling specific instances. Since this strategy is based on a single retrieval operation (Conrad, Brown, & Cashman, 1988) to which respondents probably do not have access (Ericsson & Simon, 1984), respondents may prefer this strategy and don't express their thoughts. Verbal expression, however, may give indications about the adequacy of their responses. As Houtkoop-Steenstra (2000) describes, respondents often add hedge expressions such as 'I guess' or 'probably' to their answers. A validation study by Draisma and Dijkstra (2004) shows that such linguistic indicators of uncertainty occur more frequently when responses appear to be incorrect than when they are correct.

DOI: 10.1002/acp

After a respondent's expression of uncertainty (b6), depending on her commitment to standardised interviewing (b7), an interviewer may accept the response (b9) or probe to obtain a response that the respondent is more certain about (b8). Quite likely, the latter reaction will lead to better data, although no empirical evidence is known that supports this assumption.

FORMATTING THE RESPONSE

When respondents have retrieved and judged the relevant information, they have to formulate an answer according to the response format available to them in the question. Tourangeau et al. (2000) distinguish two processes within this step. The first is mapping the answer onto the appropriate scale or response options. The second, 'editing' the response, entails that respondents adapt their answer to criteria such as consistency, social desirability, intrusiveness or politeness. Editing is less likely to be verbally expressed in the interaction and therefore will not be discussed here. With respect to response formatting, several problems may occur for which verbal evidence is likely; these are displayed in boxes c and d in Figure 3.

Mapping answers to response options is typical for survey interviews, where closed-ended questions with non-negotiable alternatives make up the majority of the questions. In ordinary conversations it is often not necessary to answer questions with that much precision. When respondents in survey interviews are asked how many days a week they watch television, they may think an answer like 'Most days' is acceptable. However, such an answer is not directly codeable by the interviewer because it does not match an exactly defined number of days (i.e. it is a *mismatch* answer). Respondents who view the survey interview as an ordinary conversation (c), may be more likely to give mismatch answers (c2) than those who view it as a formal conversation.

It is also possible that interviewers increase the probability of mismatch answers. For example, interviewers may differ with respect to the conversational character of the interview they evoke. They may motivate respondents to elaborate, hoping to make the interview a pleasant experience. This will confirm the respondents' idea that exact answers are not required.

After a mismatch answer like 'most days' in response to the question 'how many days a week do you watch television', the interviewer needs to obtain a translation of the mismatch answer into a score (the number of days). Thus, she has to probe until the respondent replies with such an exact answer (c4). According to standardisation rules, this probe should be non-directive (e.g. a probe like 'how many days would that be?', or repeating all response alternatives if available). However, to respondents such a strictly non-directive probe may imply that the interviewer was not listening. By offering only one alternative (c5) that seems to be warranted by the respondent's first answer (e.g. 'Seven days?'), the interviewer not only signals that she did pay attention to the respondent's utterance, but also makes the respondent's job easier. This is a directive probe in that it may bias respondents to accept the proposed answer. Indeed, mismatch answers appear to be the most important cause of such directive probes (Smit, 1995).

Although offering only one or a few alternatives may improve rapport, the interviewer may not always be able to accurately determine the relevant range of answers, as Excerpt 3 illustrates. From the respondent's answer in Line 2 the interviewer infers that 'once a month' might be an appropriate answer (Line 3). Such a suggestion is likely to influence the

DOI: 10.1002/acp

respondent's answer. Smit, Dijkstra, and van der Zouwen (1997) confirmed in an experimental study that respondents often accept interviewers' suggestions and found that response distributions differed between more and less suggestive probing conditions. Therefore, suggestive probing should be considered a serious problem. In Excerpt 3 the respondent happens to deny the interviewer's suggestion, but gives another mismatch answer (Line 4). In line 5 the interviewer suggests another response alternative, and the respondent gives yet another mismatch answer (Line 6). This last mismatch answer can be viewed as one that is inadequate only in a formal sense (Moore & Maynard, 2002, p. 302–303), because it unequivocally refers to one response alternative (2), unlike the substantive mismatch answers in Lines 2 and 4.

Excerpt 3: problems in formatting the response

1. I: Yes how often do you use uh do you use the Internet, E-mail or uh the World Wide Web?
2. R: Uhmmm, well I just said I just started, so that is not too often, no
3. I: Not often, just once a month?
4. R: No no no, it must be more often
5. I: Multiple times a week
6. R: It must be twice a week
7. I: Multiple times a week

This Excerpt concerns a Q–A sequence taken from the European Social Survey (see also Ongena, 2003) and translated from Dutch. The response options (on a show card), were: 1: Every day, 2: Multiple times a week, 3: Once a week, 4: Multiple times a month, 5: Once a month, 6: Less often, 7: Never, 0 (Don't know).

Interviewers may also avoid probing all together and use their own interpretation of the respondent's mismatch answer to arrive at the appropriate alternative (much as with reports). For example, the interviewer in Excerpt 3 could have just scored '1 month' right after the respondents' first answer in Line 2, without any probing. Dijkstra and van der Zouwen (1988) labelled this kind of behaviour with the term 'choosing'. In this way the interviewer, rather than the respondent, decides which response alternative is appropriate.

Moore and Maynard (2002) conclude that interviewers should be instructed not to probe after formally inadequate mismatch answers. While this makes sense in theory, we suspect that in practice interviewers will not always be able to distinguish such formally inadequate answers from substantive ones for which probing is necessary. Furthermore, a formally inadequate mismatch answer may be a means for respondents to communicate they that have not firmly decided on one of the response alternatives; in Excerpt 2, the respondent may not be satisfied with 'multiple times a week' as a translation of 'twice a week'.

Rather than accepting mismatch answers as Moore and Maynard (2002) recommend it is more efficient to prevent them in the first place. Ongena, Dijkstra, and Draisma (2005; see also Ongena, 2005) showed conversational alternatives (i.e. words common in ordinary conversations) yielded fewer mismatch answers than questions with formal response alternatives (i.e. non-common words). Thus, using alternatives that match the way people respond in ordinary conversation, prevents (or at least makes less frequent) the

DOI: 10.1002/acp

occurrence of mismatch answers, which in turn reduces the need for the interviewer to probe or infer the answer.

More talkative respondents may also elaborate their answer (c6), and the interviewer may subsequently, depending on time constraints (c7), cut-off respondent's stories (c9), instead of awaiting the respondent's full elaboration (c8).

FINALISING THE RESPONSE

When comprehension, retrieval, judgement and formatting yield no apparent problems, the respondent will formulate an initial 'adequate' answer (d). However, an answer that is produced without verbal evidence of problems could be problematic. For example such an answer could be a 'don't know' or a refusal (e). In this case, the interviewer may motivate the respondent to give a substantive answer (e2), or just record 'don't know' or 'refusal' (e3). Again, no research has been conducted to establish the frequency of occurrence of motivating and probing after don't know answers, and the effectiveness thereof. Finally, if the answer is a substantive one, the interviewer will establish its adequacy (f), for example its consistency with earlier answers. If the response is adequate, the interviewer will record the answer (f2), otherwise the interviewer may probe (f1). As Moore and Maynard (2002) state, standardisation rules do not allow interviewers to probe or seek resolution when they detect inconsistencies. However, with computer assisted interviewing, checks for consistency are automatically performed, and this forces the interviewer to confront the respondent with inconsistencies. Patterns of such correction-interactions and the consequences for data quality have not been studied yet.

CONCLUSION

The model presented in this paper indicates what aspects of survey interview interaction might be problematic for the quality of the data obtained and are worth studying. Cognitive processes are often accompanied by verbal expressions, and we assume that verbal behaviours can be used as indicators of cognitive problems. Problematic verbal behaviours may have their roots in different cognitive phases and thus may warrant different interviewer behaviour.

We don't pretend to have presented a model that can be tested at once. Rather, our model serves the purpose of generating new hypotheses to fill in the gaps with respect to insufficient empirical knowledge. Thus we devote the remainder of the paper to those research issues raised by the modelling effort that we believe to be most promising theoretically and practically.

Interviewer's question formulation

We discussed that the interviewer may have several reasons to reword questions. Although some studies exist that show relations between rewording and response accuracy, we do not know when an interviewer's rewording is likely to increase accuracy and whether the effects on response accuracy are related to the cause of rewording (politeness norms, reducing boredom or preventing problems in understanding).

DOI: 10.1002/acp

Allowing interviewers to change question wordings into recipient designed questions might solve interactional problems and may motivate respondents to give more accurate responses. However, recipient designed questions may falsely signal to respondents that general (i.e. imprecise or mismatch) answers are sufficient. Since interviewers are likely to actually infer answers from mismatch answers, recipient designed questions may decrease data quality. To test if recipient designed questions improve the data quality through establishment of rapport, or worsen of data quality through evoking mismatch answers, an experiment could be conducted in which half of the interviewers are given the freedom to change question wordings, whereas the other half should be instructed to always read questions exactly as worded. Dependent variables in such an experiment would not only be response validity (as established through for instance record checks), but also the occurrence of mismatch answers in the interaction and respondent's evaluations of the interview atmosphere.

Furthermore, we do not know whether interviewers are able to efficiently tailor question wordings to specific respondents. Tailoring is necessarily based on assumptions about the respondent. For example, interviewers might assume that it is not necessary to explain medical terms when they happen to interview a doctor. Tailoring may be more effective (i.e. result in higher response validity and fewer requests for clarification from respondents) when it is based on more versus less justifiable assumptions. For example, when it becomes apparent that respondents are doctors (based on early questions) it is easier to justify the assumption that they understand medical terms without explanation (tailoring) than when respondents profession turns out to be non-medical. To test this idea, tailoring could be compared across respondent professional groups.

Studies could also be focused on the use of CAI software to automatically adapt question wordings to respondents' specific situations.

Interpretation of a question

Providing clarification, either deliberately or in response to respondents' requests, does improve the quality of responses (Conrad & Schober, 2000). It is not known, however, when respondents are likely to ask for clarification, or give implicit requests. We hypothesise that exposure to WIMTY's ('Whatever it means to you' neutral probes) could discourage respondents from explicitly requesting clarification, whereas obtaining successful clarification will stimulate such requests. Such factors can be manipulated experimentally by instructing some interviewers to administer WIMTY probes and others not to. We would expect requests for clarification to drop off sharply as soon as they are met with WIMTY probes.

Assuming interviewers can provide substantive responses to requests for clarification. A real concern is that this will substantially inflate the interview duration. Thus unnecessary clarification ought to be minimised. Thus more knowledge is needed on when paralinguistic and visual cues that could signal comprehension difficulty actually do reflect respondent problems of this sort. Similarly, more knowledge is needed about the ability of interviewers to recognise cues signifying difficulties in understanding survey concepts. Finally it is important both theoretically and practically to determine if interviewers can be trained to accurately recognise respondent cues of comprehension difficulty.

DOI: 10.1002/acp

Retrieval/judgement

As Means et al.'s (1991) study suggested, allowing a narrative style in answering may improve data quality through provision of recall cues. However, we do not know to what extent respondents spontaneously use narratives or what their reasons are for doing so. Respondents who spontaneously use narratives may have higher task involvement, or they may have more difficulties in retrieving information and use narratives as a means to seek help form interviewers. Thus, we need to measure respondents' task involvement and cognitive capacity, and relate that to the occurrence of spontaneous narratives. We also lack quantitative information of how interviewers deal with retrieval and judgement problems. How often and when do interviewers actually interrupt respondent's narratives and how does this affect data quality?

Formatting

The more respondents use the exact wording of the response alternatives, the less the interviewer has to intervene. When response alternatives are worded according to conversational conventions, mismatch answers are much less likely to occur (Ongena et al., 2005). We have good reason to believe that this affects the quality of the data positively, but this assumption needs empirical testing.

In the Ongena et al. (2005) study, formally worded questions somewhat reduced the number of mismatch answers, but a stronger manipulation of responses alternatives could well increase the size of the effect. Hence this phenomenon needs another empirical study that compares the effect of conversational and formal response alternatives together with other factors that may affect the number of mismatch answers such as a more formal or personal style of interviewing (van der Zouwen, Dijkstra, & Smit, 1991).

Further development of the model

To test and elaborate the models of interviewer and respondent interaction as presented here, the interaction between interviewer and respondent should be more systematically studied. For such studies, a coding scheme must be detailed enough to detect problems and their interactional causes on one hand, but still produce reliable codes. A scheme that fulfils such criteria is Dijkstra's (1999) coding scheme. Using programmes like Dijkstra's Sequence Viewer (Dijkstra, 2002) it is possible to discover patterns of codes that enable researchers to understand interviewer–respondent interaction—a fundamentally qualitative phenomenon—and characterise it quantitatively.

Finally, future studies of interaction need to give greater weight to the quality of answers. It is not enough to characterise the interaction and leave it there. To really advance survey methods by studying interaction, researchers must be able to point to the impact of particular types of interaction on the accuracy of responses. This will enable further development of the model presented here and, more generally, allows methodologists to make recommendations for better survey practice.

ACKNOWLEDGEMENTS

The authors would like to thank the three anonymous reviewers for their comments on an earlier version of this paper, and Harlan Sayles for correcting the grammar.

DOI: 10.1002/acp

REFERENCES

Billiet, J., & Loosveldt, G. (1988). Improvement of the quality of responses to factual survey questions by interviewer training. *Public Opinion Quarterly, 52*, 190–211.

Bradburn, N. M., & Sudman, S. (1980). *Improving interview method and questionnaire design*. San Francisco: Jossey-Bass Publishers.

Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.

Cahalan, M., Mitchell, S., Gray, L., Chen, S., & Tsapogas, J. (1994). Recorded interview behavior coding study: National survey of recent college graduates. *Proceedings of the ASA Section on Survey Research Method*.

Cannell, C. F., Fowler, F. J., & Marquis, K. H. (1968). The influence of interviewer and respondent psychological and behavioral variables on the reporting of household interviews. *Vital and Health Statistics, Series 2, 26*, 1–65.

Conrad, F. G., Brown, N. R., & Cashman, E. R. (1988). Strategies for estimating behavioural frequency in survey interviews. *Memory, 6*, 339–366.

Conrad, F. G., & Schober, M. F. (2000). Clarifying question meaning in a household telephone survey. *Public Opinion Quarterly, 64*, 1–28.

Dijkstra, W. (1999). A new method for studying verbal interactions in survey interviews. *Journal of Official Statistics, 15*, 67–85.

Dijkstra, W. (2002). Transcribing, coding, and analyzing verbal interactions in survey interviews. In D. W. Maynard, H. Houtkoop-Steenstra, N. C. Schaeffer, & J. van der Zouwen (Eds.), *Standardization and tacit knowledge: Interaction and practice in the survey interview* (pp. 401–425). New York: Wiley.

Dijkstra, W., & van der Zouwen, J. (1988). Types of inadequate interviewer behaviour in survey-interviews. In W. E. Saris, & I. N. Gallhofer (Eds.), *Sociometric research: Data collection and scaling* (Vol. 1, pp. 24–35). London: MacMillan.

Draisma, S., & Dijkstra, W. (2004). Response latency and (para)linguistic expressions as indicators of response error. In S. Presser, J. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 131–147). New York: Wiley.

Dykema, J., Lepkowski, J. M., & Blixt, S. (1997). The effect of interviewer and respondent behavior on data quality: Analysis of interaction coding in a validation study. In L. Lyberg, P. Biemer, M. Collins, E. De Leeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.), *Survey measurement and process quality* (pp. 287–310). New York: John Wiley & Sons Inc.

Ericsson, K. A., & Simon, H. A. (1984). *Verbal reports as data*. Cambridge, MA: Bradford Books/MIT Press.

Fowler, F. J., & Mangione, T. W. (1990). *Standardized survey interviewing: Minimizing interviewer-related error.* Newbury Park, CA: Sage.

Hansen, S. E., & Couper, M. P. (2004). Usability testing to evaluate computer-assisted instruments. In S. Presser, J. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 337–360). New York: Wiley.

Houtkoop-Steenstra, H. (2000). *Interaction in the standardized survey interview: The living questionnaire*. Cambridge: Cambridge University Press.

Hyman, H., Cobb, W. J., Feldman, J., Hart, C. W., & Stember, C. (1954). *Interviewing in social research*. Chicago: University of Chicago Press.

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology, 50*, 537–567.

Marquis, K. H., & Cannell, C. F. (1969). *A study of interviewer-respondent interaction in the urban employment survey.* Ann Arbor: Mich. Survey Research Center, University of Michigan.

Mathiowetz, N. A., & Cannel, C. F. (1980). Coding interviewer behavior as a method of evaluating performance. In *Proceedings of the section of survey research methods* (pp. 525–528). Alexandria, VA: American Statistical Association.

Means, B., Swan, G. E., Jobe, J. B., & Esposito, J. L. (1991). An alternative approach to obtaining personal history data. In P. Biemer, R. M. Grovers, L. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 167–183). New York: Wiley.

Moore, R. J. (2004). Managing troubles in answering survey questions: Respondents' uses of projective reporting. *Social Psychology Quarterly, 67*, 50–69.

DOI: 10.1002/acp

Moore, R. J., & Maynard, D. W. (2002). Achieving understanding in the standardized survey interview: Repair sequences. In D. W. Maynard, H. Houtkoop-Steenstra, N. C. Schaeffer, & J. van der Zouwen (Eds.), *Standardization and tacit knowledge: Interaction and practice in the survey interview* (pp. 281–311). New York: Wiley.

Ongena, Y. P. (2003). Pre-testing the ESS-questionnaire using interaction analysis. *Paper Presented at the European Social Survey CCT Meeting*, Sociaal en Cultureel Planbureau., Den Haag.

Ongena, Y. P. (2005). *Interviewer and Respondent Interaction in Survey Interviews*. Doctoral Dissertation. Amserdam: Vrije Universiteit https://dare.ubvu.vu.nl/bitstream/1871/9842/1/6353.pdf.

Ongena, Y. P., & Dijkstra, W. (2006). Methods of behavior coding of survey interviews. *Journal of Official Statistics, 22*, 419–451.

Ongena, Y. P., Dijkstra, W., & Draisma, S. (2005). Conversational and formal questions in Survey Interviews. In *Proceedings of the RC33 6th International Conference On Social Science Methodology*: Amsterdam 2004. CD-ROM Barbara Budrich Publishers. ISBN 3-908394-44-33.

Pierzchala, M., & Manners, T., II (1998). Producing CAI instruments for a program of surveys. In M. P. Couper, R. P. Baker, J. Bethlehem, C. Z. F. Clark, J. Martin, W. Nicholls, & J. M. O'Reilly (Eds.), *Computer assissted survey information collection* (pp. 125–145). New York: Wiley.

Sander, J. E., Conrad, F. G., Mullin, P. A., & Herrmann, D. J. (1992). Cognitive modelling of the survey interview. *Proceedings of the Joint Statistical Meetings, Survey Research Methods Section*. 818–823.

Schaeffer, N. C., & Maynard, D. W. (2002). Occasions for intervention: Interactional resources for comprehension in standardized survey interviews. In D. W. Maynard, H. Houtkoop-Steenstra, N. C. Schaeffer, & J. van der Zouwen (Eds.), *Standardization and tacit knowledge: Interaction and practice in the survey interview* (pp. 261–280). New York: John Wiley & Sons, Inc.

Schober, M. F. (1999). Making sense of questions: An interactional approach. In M. G. Sirken, D. J. Herrman, S. Schechter, N. Schwarz, J. M. Tanur, & R. Tourangeau (Eds.), *Cognition and survey research* (pp. 77–93). New York: John Wiley & Sons Inc.

Schober, M. F., & Bloom, J. E. (2004). Discourse cues that respondents have misunderstood survey questions. *Discourse Processes, 38*, 287–308.

Schober, M. F., & Conrad, F. G. (1997). Does conversational interviewing reduce survey measurement error? *Public Opinion Quarterly, 61*, 576–602.

Schober, M. F., & Conrad, F. G. (2002). A collaborative view of standardized survey interviews. In D. W. Maynard, H. Houtkoop-Steenstra, N. C. Schaeffer, & J. van der Zouwen (Eds.), *Standardization and tacit knowledge: Interaction and practice in the survey interview* (pp. 67–94). New York: Wiley.

Schober, M. F., Conrad, F. G., & Fricker, S. S. (2004). Misunderstanding standardized language in research interviews. *Applied Cognitive Psychology, 18*, 169–188.

Smit, J. H. (1995). *Suggestieve vragen in survey-interviews. Vóórkomen, oorzaken en gevolgen*. Amsterdam: Academisch proefschrift, Vrije Universiteit.

Smit, J. H., Dijkstra, W., & van der Zouwen, J. (1997). Suggestive interviewer behaviour in surveys: An experimental study. *Journal of Official Statistics, 13*, 19–28.

Smit, E. G., & Neyens, P. C. (2000). Segmentation based on affinity for advertising. *Journal of Advertising Research, 40*, 35–43.

Sperry, S., Edwards, B., Dulaney, R., & Potter, D. E. B., II (1998). Evaluating interviewer use of CAPI navigation features. In M. P. Couper, R. P. Baker, J. Bethlehem, C. Z. F. Clark, J. Martin, W. Nicholls, & J. M. O'Reilly (Eds.), *Computer assissted survey information collection* (pp. 351–365). New York: Wiley.

Suchman, L., & Jordan, B. (1990). Interactional troubles in face-to-face survey interviews. *Journal of the American Statistical Association, 85*, 232–253.

Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology.* San Francisco: Jossey-Bass.

Tourangeau, R., Rips, L. C., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.

van der Zouwen, J., Dijkstra, W., & Smit, J. H. (1991). Studying respondent-interviewer interaction: The relationship between interviewing style, interviewer behavior, and response behavior. In P. Biemer, R. M. Groves, L. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 419–437). New York: Wiley.

DOI: 10.1002/acp